Environmental Health Perspectives
Vol. 63, pp. 235-239, 1985

Analysis of Data in Square Contingency Tables with Ordered Categories Using the Conditional Symmetry Model and its Decomposed Models

by Sadao Tomizawa*

For the analysis of square contingency tables with ordered categories, three kinds of decompositions for the conditional symmetry model derived by Tomizawa are simply described. Using the conditional symmetry model and its decomposed models, this paper analyzes the data of unaided distance vision of women in Britain first analyzed by Stuart, the data of unaided distance vision of students in a university in Japan, and the data of unaided distance vision of pupils at elementary schools at a city in Tokyo.

Introduction

Table 1 is constructed from the data of the unaided distance vision of 7477 women aged 30–39 employed in Royal Ordnance factories in Britain from 1943 to 1946. The data in Table 1 were first analyzed by Stuart (*1*,*2*). Table 2 is constructed from the data of the unaided distance vision of 4746 students aged 18 to about 25, including about 10% of the women of the Faculty of Science and Technology, Science University of Tokyo in Japan examined in April, 1982. Table 3 is constructed from the data of the unaided distance vision of 3168 pupils aged 6–12, including about half the girls at elementary schools in Tokyo, Japan examined in June 1984. In Tables 1, 2, and 3 the row variable is the right eye grade and the column variable is the left eye grade with the categories ordered from the lowest grade (1) to the highest grade (4).

To the data of Tables 1, 2, and 3 it is reasonable to apply models of various kinds of symmetry instead of the statistical independence model. To analyze the data of square contingency tables, the models of symmetry, quasisymmetry and marginal homogeneity are described, for example, in Bishop, Fienberg, and Holland (*3*), Caussinus (*4*), and Stuart (*2*). Caussinus (*4*) also noted that the symmetry model holds if and only if both quasisymmetry model and marginal homogeneity model hold. McCullagh (*5*) proposed a conditional symmetry model which is an extension of the symmetry model. Tomizawa (*6*) derived three kinds of decompositions for the conditional symmetry model. Wall and Lienert (*7*) proposed a point-symmetry model in J-dimensional contingency cubes. Tomizawa (*8*,*9*) proposed models of various kinds of point symmetry in two-dimensional contingency tables and gave their decompositions.

In this paper we analyze the data in Tables 1, 2, and 3 using the conditional symmetry model and its decomposed models.

Table 1. Unaided distance vision of 7477 women aged 30–39 employed in Royal Ordnance factories from 1943 to 1946.

	Left eye grade				
Right eye grade	Lowest (1)	Second (2)	Third (3)	Highest (4)	Total
Lowest (1)	492	179	82	36	789
Second (2)	205	1772	362	117	2456
Third (3)	78	432	1512	234	2256
Highest (4)	66	124	266	1520	1976
Total	841	2507	2222	1907	7477

Table 2. Unaided distance vision of 4746 students aged 18 to about 25 including about 10% women in Faculty of Science and Technology, Science University of Tokyo in Japan examined in April 1982.

	Left eye grade				
Right eye grade	Lowest (1)	Second (2)	Third (3)	Highest (4)	Total
Lowest (1)	1429	249	25	20	1723
Second (2)	185	660	124	64	1033
Third (3)	23	114	221	149	507
Highest (4)	22	40	130	1291	1483
Total	1659	1063	500	1524	4746

*Department of Information Sciences, Faculty of Science and Technology, Science University of Tokyo, Noda City, Chiba 278 Japan.

Table 3. Unaided distance vision of 3168 pupils aged 6–12 including about half girls at elementary schools in Tokyo examined in June 1984.

	Left eye grade				
Right eye grade	Lowest (1)	Second (2)	Third (3)	Highest (4)	Total
Lowest (1)	92	16	7	12	127
Second (2)	15	75	42	10	142
Third (3)	5	33	138	96	272
Highest (4)	10	21	126	2470	2627
Total	122	145	313	2588	3168

Models and Decompositions

Consider the square $a \times a$ contingency table with row variate denoted by X_1 and column variate denoted by X_2. Let p_{ij} denote the probability in the cell in row i and column j for $1 \leq i,j \leq a$.

The models of symmetry, quasisymmetry and marginal homogeneity are defined as follows:

$$H_S: \ p_{ij} = p_{ji} \qquad \text{for } 1 \leq i < j \leq a$$
$$H_Q: \ p_{ij} = a_i b_j d_{ij} \qquad \text{for } 1 \leq i,j \leq a$$

where $a_i > 0$, $b_j > 0$, $d_{ij} > 0$, and $d_{ij} = d_{ji}$;

$$H_M: \ p_{i\cdot} = p_{\cdot i} \qquad \text{for } 1 \leq i \leq a$$

where

$$p_{i\cdot} = \sum_{l=1}^{a} p_{il}$$

and

$$p_{\cdot i} = \sum_{l=1}^{a} p_{li}$$

McCullagh's (5) conditional symmetry model is defined by

$$H_S{*}: p_{ij} = \begin{cases} \theta\,\Phi_{ij} & (i < j) \\ \Phi_{ii} & (i = j) \\ (2-\theta)\,\Phi_{ij} & (i > j) \end{cases}$$

where $\Phi_{ij} = \Phi_{ji}$ and $\Sigma\Sigma\Phi_{ij} = 1$. This model can be also expressed by a log-linear model for the p_{ij} in Tomizawa (6).

We next define the extended quasisymmetry model by

$$H_Q{*}: p_{ij} = a_i b_j d_{ij} \qquad \text{for } 1 \leq i,j \leq a$$

where $a_i > 0$, $b_j > 0$, $d_{ij} > 0$

$$d_{ij} = \gamma d_{ji} \quad \text{for } 1 \leq i < j \leq a$$

$$\prod_{l=1}^{a} a_l = \prod_{l=1}^{a} b_l = 1$$

$$\prod_{l=1}^{a} (d_{il}/d_{lj}) = \gamma^{a+1-(i+j)}$$

and where the parameter γ is unspecified. This model is also equivalent to the model

$$H_Q{*}: p_{ij}p_{jk}p_{ki} = \gamma p_{ji}p_{kj}p_{ik} \quad \text{for } 1 \leq i < j < k \leq a$$

where the parameter γ is unspecified. Model H_Q^* is also expressed by a log-linear model for the p_{ij} in Tomizawa (6). A special case of model H_Q^* obtained by putting $\gamma = 1$ is the quasisymmetry model.

We introduce two kinds of modified marginal homogeneity models as follows:

$$H_{M1}{*}: p_{i\cdot}^{+} = \delta p_{\cdot i}^{-} \qquad \text{for } 1 \leq i \leq a-1$$

where the parameter δ is unspecified and where

$$p_{i\cdot}^{+} = \Pr(X_1 = i, X_1 < X_2) = \sum_{l=i+1}^{a} p_{il}$$

$$p_{\cdot i}^{-} = \Pr(X_2 = i, X_1 > X_2) = \sum_{l=i+1}^{a} p_{li}$$

$$H_{M2}{*}: \ p_{\cdot i}^{+} = \delta p_{i\cdot}^{-} \qquad \text{for } 2 \leq i \leq a$$

where the parameter δ is unspecified and where

$$p_{\cdot i}^{+} = \Pr(X_2 = i, X_1 < X_2) = \sum_{l=1}^{i-1} p_{li}$$

$$p_{i\cdot}^{-} = \Pr(X_1 = i, X_1 > X_2) = \sum_{l=1}^{i-1} p_{il}$$

We define the extended marginal homogeneity model by

$$H_{M3}{*}: \ p_{i\cdot}^{(\delta)} = p_{\cdot i}^{(\delta)} \qquad \text{for } 1 \leq i \leq a$$

where the parameter δ is unspecified and where

$$p_{i\cdot}^{(\delta)} = \delta\, p_{i\cdot}^{-} + p_{ii} + p_{i\cdot}^{+} \text{ and } p_{\cdot i}^{(\delta)} = p_{\cdot i}^{+} + p_{ii} + \delta\, p_{\cdot i}^{-}.$$

Model H_{M3}^* indicates that the row marginal totals summed by multiplying the probabilities for the cells in the lower-left triangle of the $a \times a$ table by a common weight δ are equal to the column marginal totals summed by the same way. Model H^*_{M3} also has the property that under H^*_{M3} the parameter $\delta \geq 1$ is equivalent to $\Pr(X_1 \leq i) \geq \Pr(X_2 \leq i)$ for every i. A special case of model H^*_{M3} obtained by putting $\delta = 1$ is the marginal homogeneity model.

Finally let

$$R = \sum_{i<j<k} \{p_{ij}p_{jk}p_{ki}/(p_{ji}p_{kj}p_{ik})\} / \binom{a}{3}$$

and introduce three kinds of average models as follows:

$$\mathrm{H}^*_{R1} : \ R = \Delta_1$$

where $$\Delta_1 = \sum_{i=1}^{a-1} (p^+_{i\cdot}/p^-_{\cdot i})/(a - 1)$$

$$\mathrm{H}^*_{R2} : \ R = \Delta_2$$

where $$\Delta_2 = \sum_{i=2}^{a} (p^+_{\cdot i}/p^-_{i\cdot})/(a - 1)$$

$$\mathrm{H}^*_{R3} : \ R = \Delta_3$$

where $$\Delta_3 = (p^+_{1\cdot}/p^-_{\cdot 1} + p^+_{\cdot a}/p^-_{a\cdot})/2$$

Here R indicates the average of ratio $p_{ij}p_{jk}p_{ki}/(p_{ji}p_{kj}p_{ik})$ for $1 \leq i < j < k \leq a$ in the case that the ratio parameter γ in H^*_Q changes according to each ratio, and Δ_l indicates the average of $p^+_{i\cdot}/p^-_{\cdot i}$ for $1 \leq i \leq a - 1$ in the case that the ratio parameter δ in H^*_{M1} changes according to each ratio, and also Δ_2 is interpreted similarly, and Δ_3 indicates the average of two ratios in the case that the ratio $p^+_{1\cdot}/p^-_{\cdot 1}$ is not always equal to the ratio $p^+_{\cdot a}/p^-_{a\cdot}$. Therefore model H^*_{Rl} for $l = 1, 2$, and 3 indicates the equilibrium of two kinds of averages R and Δ_l.

We get the decompositions for model H^*_S as follows.

THEOREM: for $l = 1, 2$, and 3, model H^*_S holds if and only if all models H^*_Q, H^*_{Ml} and H^*_{Rl} hold.

The proof of this theorem is given by Tomizawa (*6*).

We denote special cases of models $H^*_{Ml}(l = 1,2)$ obtained by putting $\delta = 1$ by H_{Ml} $(l = 1,2)$. Then we get new decompositions for model H_S as follows:

COROLLARY: for $l = 1$ and 2, model H_S holds if and only if both models H_Q and H_{Ml} hold.

Degrees of Freedom, Estimation, and Test

Let x_{ij} denote the observed frequency in the ith row and jth column of the $a \times a$ contingency table $(1 \leq i,j \leq a)$ where $\sum\sum x_{ij} = N$, and let m_{ij} denote the corresponding expected frequency under some model. We assume here that a multinomial distribution applies to the $a \times a$ table.

The degrees of freedom for models H^*_S, H^*_Q, H^*_{Ml}, H^*_{Rl} $(l = 1,2,3)$ and $H_{Ml}(l = 1,2)$ are $(a - 2)(a + 1)/2$, $a(a - 3)/2$, $a - 2$, 1, and $a - 1$, respectively.

The maximum likelihood estimates of m_{ij} under model H^*_Q can be sought by the iterative procedure in Tomizawa (*6*) or by the following iterative procedure: as the $(k+1)$-th step

$$\hat{m}_{ij}^{(k+1)} = \hat{m}_{ij}^{(k)} \left[\frac{x_{i\cdot}x_{\cdot j}\{x_{ij} + x_{ji}\}}{\hat{m}_{i\cdot}^{(k)}\hat{m}_{\cdot j}^{(k)}\{\hat{m}_{ij}^{(k)} + \hat{m}_{ji}^{(k)}\}} \right]^{1/4} \times \left[\frac{D}{D^{(k)}}\right]^{d_{ij}/4} \left[\frac{E}{E^{(k)}}\right]^{(1-d_{ij}+)/4}$$

where the initial values are $\hat{m}_{ij}^{(0)} = 1$ for $1 < i,j \leq a$ and where

$$x_{i\cdot} = \sum_{l=1}^{a} x_{il}$$

$$x_{\cdot j} = \sum_{l=1}^{a} x_{lj}$$

$$\hat{m}_{i\cdot}^{(k)} = \sum_{l=1}^{a} \hat{m}_{il}^{(k)}$$

$$\hat{m}_{\cdot j}^{(k)} = \sum_{l=1}^{a} \hat{m}_{lj}^{(k)}$$

$$d_{ij} = \{a + 1 - (j - i)\}/(a - 2) \qquad (i < j)$$

$$d_{ij} = 1/2 \qquad (i \geq j)$$

$$D = (a - 2)N + \sum_{i<j} x_{ij}\{a - 2(j - i)\}$$

$$E = 2(a - 2)N - D$$

$$D^{(k)} = (a - 2) \sum_{i,j} \hat{m}_{ij}^{(k)} + \sum_{i<j} \hat{m}_{ij}^{(k)} \{a - 2(j - 2)\}$$

$$E^{(k)} = 2(a - 2) \sum_{i,j} \hat{m}_{ij}^{(k)} - D^{(k)}$$

Thus the goodness of fit of models H^*_S, H^*_Q, H^*_{Ml} and $H^*_{Ml}(l = 1,2)$ can be tested by the Pearson's or the likelihood-ratio chi-squared statistics. The goodness of fit of model H^*_{M3} can be tested by test statistic χ^{*2}_{M3} in Tomizawa (*6*).

Table 4. Chi-square for symmetry models applied to the data in Table 1.

Symmetry models	Degrees of freedom	Likelihood-ratio chi-square	Pearson's chi-square
H_S	6	19.25	19.11
H_Q	3	7.27	7.26
H_{M1}	3	11.97	11.96
H_{M2}	3	11.99	11.97
H_S^*	5	7.35	7.26
H_Q^*	2	6.82	6.78
H_{M1}^*	2	0.08	0.08
H_{M2}^*	2	0.09	0.09

Table 5. Chi-square for symmetry models applied to the data in Table 2.

Symmetry models	Degrees of freedom	Likelihood-ratio chi-square	Pearson's chi-square
H_S	6	16.95	16.87
H_Q	3	5.71	5.78
H_{M1}	3	12.52	12.49
H_{M2}	3	13.94	13.90
H_S^*	5	4.98	4.97
H_Q^*	2	4.41	4.39
H_{M1}^*	2	0.54	0.54
H_{M2}^*	2	1.96	1.96

Table 6. Chi-square for symmetry models applied to the data in Table 3.

Symmetry models	Degrees of freedom	Likelihood-ratio chi-square	Pearson's chi-square
H_S	6	9.69	9.58
H_Q	3	2.81	2.75
H_{M1}	3	4.49	4.48
H_{M2}	3	6.98	6.95
H_S^*	5	7.83	7.77
H_Q^*	2	2.61	2.57
H_{M1}^*	2	2.63	2.63
H_{M2}^*	2	5.12	5.12

Analysis of Table 1

Table 4 presents the likelihood-ratio and the Pearson's chi-squared statistics obtained by applying the models introduced in the previous section to the data in Table 1. The value of test statistic Q in Stuart (*2*) for testing the goodness of fit of model H_M is 11.96 with 3 degrees of freedom. The value of test statistic χ_{M3}^{*2} in Tomizawa (*6*) for testing the goodness of fit of the extended marginal homogeneity model H_{M3}^* is 0.005 with 2 degrees of freedom. From these values and Table 4, none of models H_{M1}, H_{M2}, and H_M fits the data well, but all of models H_{M1}^*, H_{M2}^*, and H_{M3}^* fit the data very well. Moreover the maximum likelihood estimates of δ under models H_{M1}^* and H_{M2}^* are 0.863. Since this value is less than one, we can say that the left eye is worse than the right eye. Also the values of chi-square under model H_Q^* lie between the upper 5% and 1% tail values of the χ^2 distribution with 2 degrees of freedom. Under model H_Q^* the estimated value of γ obtained by maximum likelihood is 0.929. Also model H_S does not fit the data well, and thus the left eye is not symmetric to the right eye. But model H_S fits adequately and the estimated values of m_{ij}/m_{ji} for $1 \leq i < j \leq 4$ under model H_S^* are 0.863. Since this value is less than one, we can say again that the left eye is worse than the right eye.

Analysis of Table 2

Table 5 presents the likelihood-ratio and the Pearson's chi-squared statistics obtained by applying various kinds of symmetry models to the data in Table 2. The value of test statistic Q in Stuart (*2*) for testing the goodness of fit of model H_M is 11.21 with 3 degrees of freedom. This value lies between the upper 5% and 1% tail values of the χ^2 distribution with 3 degrees of freedom. The value of test statistic χ_{M3}^{*2} in Tomizawa (*6*) for testing the goodness of fit of the extended marginal homogeneity model H_{M3}^* is 0.56 with 2 degrees of freedom. From this value and Table 5, neither model H_{M1} nor model H_{M2} fits the data well but all of models H_{M1}^*, H_{M2}^*, and H_{M3}^* fit the data very well. Moreover the maximum likelihood estimates of δ under models H_{M1}^* and H_{M2}^* are 1.228, and since this value is greater than one, this value indicates that the left eye is better than the right eye. Both models H_Q and H_Q^* also fit adequately, and under model H_Q^* the estimated value of γ obtained by maximum likelihood is 1.208. Also, since model H_S does not fit the data well, the left eye is not symmetric to the right eye. But model H_S^* fits the data well, and the estimated values of m_{ij}/m_{ji} for $1 \leq i < j \leq 4$ under model H_S^* are 1.228. Since this value is greater than one, we can say again that the left eye is better than the right eye.

Analysis of Table 3

Table 6 presents the likelihood-ratio and the Pearson's chi-squared statistics obtained by applying various kinds of symmetry models to the data in Table 3. The value of test statistic Q in Stuart (*2*) for testing the goodness of fit of model H_M is 6.85 with 3 degrees of freedom, and the value of test statistic χ_{M3}^{*2} in Tomizawa (*6*) for testing the goodness of fit of the extended marginal homogeneity model H_{M3}^* is 4.16 with 2 degrees of freedom. From these values and Table 6, all models fit the data well. Moreover under model H_Q^* the estimated value of γ obtained by maximum likelihood is 1.136 and the maximum likelihood estimates of δ under models H_{M1}^* and H_{M2}^* are 0.871. We may consider these values close upon one because models H_Q, H_{M1}, and H_{M2} hold. Also the values of statistics for the goodness of fit of models H_S^* and $H_{Ml}^*(l=1, 2, 3)$ applied to Table 3 are greater than those applied to Table 1 and 2; namely, the goodness of fit of models H_S^* and $H_{Ml}^*(l=1, 2, 3)$ applied to Table 3 are not so good as those applied to Table 1 and 2. By the way, models H_S, H_{M1}, H_{M2}, and H_M applied to Table 1 and 2 did not fit the data well, but those applied to Table 3 fit the data well. Therefore, for the data in Table 3 we can say that the left eye is symmetric to the right eye in the various senses.

The author would like to express his sincere thanks to Professor K. Kunisawa, Mr. M. Jimbo, and Dr. K. Shimizu in Science University of Tokyo for their valuable advice and encouragements.

The author is grateful to Dr. A. Nakajima and Dr. K. Fujiki in Department of Ophthalmology, Juntendo University, for their valuable advice and for giving him the data in Table 3 in this paper.

He is also deeply grateful to Dr. A. Kudo and Dr. T. Yanagawa in Kyushu University who provided him with an excellent opportunity for presenting this paper at the Environmental Risk Assessment and Statistical Methods Conference which is a satellite conference in conjunction with the 12th International Biometric Conference.

REFERENCES

1. Stuart, A. The estimation and comparison of strengths of association in contingency tables. Biometrika 40: 105–110 (1953).
2. Stuart, A. A test for homogeneity of the marginal distributions in a two-way classification. Biometrika 42: 412–416 (1955).
3. Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, MA, 1975.
4. Caussinus, H. Contribution à l'analyse statistique des tableaux de corrélation. Ann. Fac. Sci. Univ. Toulouse 29: 77–182 (1965).
5. McCullagh, P. A class of parametric models for the analysis of square contingency tables with ordered categories. Biometrika 65: 413–418 (1978).
6. Tomizawa, S. Three kinds of decompositions for the conditional symmetry model in a square contingency table. J. Japan Statist. Soc. 14: 35–42 (1984).
7. Wall, K. D., and Lienert, G. A. A test for point-symmetry in J-dimensional contingency-cubes. Biometrische Z. 18: 259–264 (1976).
8. Tomizawa, S. The decompositions for point symmetry models in two-way contingency tables. Biometrical J., in press.
9. Tomizawa, S. A decomposition for the inclined point-symmetry model in a square contingency table. Biometrical J., in press.